

Energy Efficient Video Telephony over UMTS

C. Kodikara, S. Worrall, A.M. Kondoz

Centre for Communication Systems Research

University of Surrey

Guildford, Surrey, GU2 7XH, U.K.

{c.kodikara, s.worrall, a.kondoz} @eim.surrey.ac.uk

Abstract— A novel joint Unequal Error Protection (UEP) and Unequal Power Allocation (UPA) algorithm is proposed. UEP is performed through multiple radio bearers with different error correction capabilities. The perceived video quality is estimated based on the distortion calculation at the encoder. The transmission power on each bearer is controlled to maximize the expected perceived video quality for a given operating condition. The algorithm's performance is evaluated for MPEG-4 coded video transmission over a simulated UMTS network. Simulation results show that the joint UEP and UPA scheme is more energy efficient than applying them individually.

Keywords—video communication, Unequal error protection, unequal power allocation, UMTS

I. INTRODUCTION

Efficient transmission power utilization is an important design criterion in interference limited cellular systems, such as UMTS networks. System capacity is limited by the total interference experienced within the cell coverage. Thus, the optimization of power consumption for an individual user can provide an increase in system capacity, as well as in the quality of service experienced by the user.

Recently, several energy minimization techniques for wireless video applications have been proposed [1, 2]. All of these techniques are optimized to achieve a target video quality while minimizing the transmission power. In [1], joint error resilience and transmission power management at the video frame level is proposed. However, as video frame quality is variable in nature (even in an error free environment), controlling the transmission power to achieve a target frame quality at the video frame level is inaccurate and would result in poor system performance. This problem can be solved by minimizing the total consumed power within a certain period, while achieving the optimal average video quality. The method proposed in [2], employs this concept for Intra-refreshed video sequences, thus video performance is optimized for a fixed Intra-refresh period. However, this method can not be applied in conjunction with rate controlled Adaptive Intra Refresh (AIR) techniques [3], which is commonly used to produce a smoother output bit rate for transmission over a fixed bandwidth channel.

Another issue that should be considered in the design of transmit power optimization schemes is the support of network compatibility and interoperability between different networks and platforms. Transmit power is normally allocated at the physical layer for a given Transmission Time Interval (TTI)

[4]. If the power allocation algorithms are closely coupled with the video compression formatting as in [1], it is impossible to implement such an algorithm at the physical layer without modifying the entire protocol stack of the existing network.

The algorithm proposed in this paper takes these issues in to account in the design and implementation. In contrast to the method in [2], the proposed scheme can equally be applied to rate controlled AIR video sequences as well as intra-refreshed sequences. The proposed method combines a Unequal Error Protection (UEP) technique and an Unequal Power Allocation (UPA) technique to obtain energy efficient video transmission. UEP is performed with the use of two different radio bearers. At the start of every video frame, the transmission energy for different bearers is selected in such a way as to achieve the maximum expected video frame quality for an increment in the transmission power step.

The remainder of this paper is organized as follows. Section II describes the distortion modeling for rate controlled MPEG-4 coded video sequences. The proposed power optimized UEP algorithm is introduced in Section III. Section IV presents the simulation set up and the algorithm implementation for UMTS networks. Experimental results and analysis are presented in Section V. Finally, conclusions are drawn in Section VI.

II. PERFORMANCE MODELING

The bit stream syntax of MPEG-4 uses a hierarchical structure. Each video frame is partitioned into smaller rectangular regions called "Macro-Blocks (MBs)" (16×16 pixels in size). Each MB is coded either in inter-mode or intra-mode. Intra-mode MBs are transform-coded directly without applying motion compensation while inter-mode MB's use motion compensation. Consecutive MB's are grouped to form Video Packets (VP). The MPEG-4 adopted video packet format is shown in Fig. 1. Synchronization markers are used to isolate VPs from each other. Following the concept of data partitioning, data within a video packet further divided into two main parts. The motion related information for all the MB's contained in a given VP are placed in the motion part and the relevant DCT data are placed in the texture part [5].

Video Object Plane (VOP) consists of a number of VPs. The most important information the decoder needs to know prior to the decoding of compressed video data, is placed in the VOP header part. This includes the spatial dimensions of the video frame, the time stamps associated with the current frame, presentation information and the mode in which the current frame is coded.

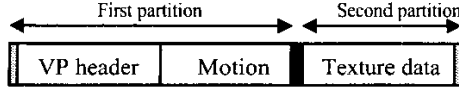


Figure 1. MPEG-4 Video Packet format.

The video information is prioritized based on the MPEG-4 data partition techniques [6]. Information within the first partition is sent over a higher priority channel. Lower priority channel is used to transmit the data within the second partition. Video packet header information is added to the beginning of second partition in order to guarantee accurate stream synchronization at the decoder. VOP header is also repeated in lower priority channel at the starting of the video frame. The video data formats in prioritized streams are shown in Fig. 2.

Video performance is modeled as a combination of quantization distortion, $E(D_{Q,pv})$ and channel distortion. Channel distortion is divided into three parts; namely spatial concealment distortion, temporal concealment distortion and distortion due to error propagation. The distortion method adopted in this paper is similar to the method proposed in [2]. However, the model in [2] is enhanced for the use of AIR and features improved error propagation estimation.

If the video frame configuration information or video packet headers are lost in the transmission, the decoding process is impossible and the data belonging to that video packet has to be discarded at the decoder. However, the error concealment tools implemented at the decoder replaces the discarded data with error-concealed data from the neighboring packets. The distortion resulting from this process is called the spatial- concealment distortion, $E(D_{s_con,pv})$, as only spatial error concealment is involved in the process. On the other hand, if the configuration information, video packet header information and motion information is received correctly but DCT information is corrupted, then the decoder only discards the corrupted DCT data and replaces them with the corresponding concealed data from the previous frame. The distortion resulting from motion compensated error concealment is called temporal concealment distortion, $E(D_{t_con,pv})$.

Errors can propagate in two ways, either in the temporal domain or in the spatial domain. Frame to frame error propagation through motion prediction and temporal concealment is called temporal domain error propagation, f_{tp} . The propagation of errors from neighboring video packets via spatial concealment is considered as spatial domain error propagation, f_{sp} .

Taking the video packet as the base unit, the expected frame quality can be written as

$$E(Q_f^j) = 10 \cdot \log(g / \sum_{i=0}^{I^j} E(D_{pv}^{i,j})) \quad (1)$$

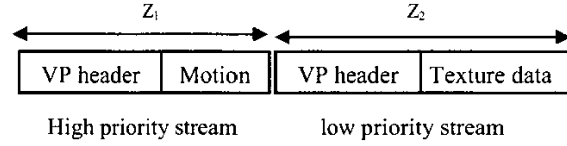


Figure 2. Video data format in prioritized streams.

where $E(Q_f^j)$ is the expected quality, $E(D_{pv}^{i,j})$ is the expected distortion of video packet and I^j is the total number of video packets in j^{th} video frame. i and j represent the i^{th} video packet of the j^{th} video frame. g is a constant defined by the dimension of the frame. $E(D_{pv}^{i,j})$ can be written as

$$E(D_{pv}^{i,j}) = E(D_{Q,pv}^{i,j}) + \rho_{u,pv}^{i,j} E(D_{s_con,pv}^{i,j}) + \rho_{d,pv}^{i,j} E(D_{t_con,pv}^{i,j}) + f_{tp}^{i,j} + f_{sp}^{i,j} \quad (2)$$

$\rho_{u,pv}^{i,j}$ denotes the probability of receiving an un-decodable video packet. This includes the corruption of VOP headers, VP headers or motion data. $\rho_{u,pv}^{i,j}$ equals the probability of finding an error in the first partition. The probability of receiving a decodable video packet but with errors, where the DCT data is corrupted but other information is received correctly is denoted by $\rho_{d,pv}^{i,j}$. This equals the probability of finding an error in the second partition but not in the first partition. Both probability terms are functions of channel bit error rate, transmission bit energy, and total background noise power.

A. Probability calculation

Assuming that the probability of receiving a VOP header with errors in the high priority channel is $\rho_{vop1}^{i,j}$, the probability of receiving data within the first partition with errors is $\rho_M^{i,j}$ and the probability of finding an error in the second partition is $\chi^{i,j}$. Then

$$\rho_{u,pv}^{i,j} = 1 - (1 - \rho_{vop1}^{i,j})(1 - \rho_M^{i,j}) \quad (3)$$

$$\rho_{d,pv}^{i,j} = (1 - \rho_{vop1}^{i,j})(1 - \rho_M^{i,j})\chi^{i,j} \quad (4)$$

If the probability of channel bit errors in channel 1 and channel 2 are denoted by ρ_{b1} and ρ_{b2} respectively, then it can be shown that

$$\rho_{vop1}^{i,j} = \sum_{c=1}^V (1 - \rho_{b1})^{c-1} \rho_{b1} = (1 - (1 - \rho_{b1})^V) \quad (5)$$

where V represents the VOP header size. Similarly,

$$\rho_M^{i,j} = 1 - (1 - \rho_{b1})^{Z_1} \quad (6)$$

$$\begin{aligned} \chi^{i,j} &= 1 - (1 - \rho_{b2})^V + (1 - \rho_{b2})^V \cdot \sum_{z=1}^{Z_2} \binom{Z_2}{z} (1 - \rho_{b2})^{Z_2-z} \rho_{b2}^z \\ &= 1 - (1 - \rho_{b2})^V \cdot (1 - \rho_{b2})^{Z_2} \end{aligned} \quad (7)$$

Z_1 and Z_2 are as defined in Fig. 2.

B. Distortion modeling

The expected distortions, $E(D_{Q,pv})$, $E(D_{s_con,pv})$ and $E(D_{t_con,pv})$ of the MBs are calculated as specified in [2]. The quantization distortion is computed by comparing the reconstructed MBs and the original MBs at the encoder. Concealment distortions are also computed in a similar manner. The transmitted video data belonging to each MB is corrupted using a noise generator located at the encoder. Corrupted data is replaced by the concealed data and data belonging to the original and concealed MB's are compared. In the case of the spatial concealment distortion calculation, only a spatial concealment algorithm is used to generate the concealed data. On the other hand, a temporal concealment algorithm alone is used to conceal the corrupted data for the temporal concealment distortion calculation. The correct reception of neighboring video packets and reference video frames is assumed in the calculations.

C. Propagation loss modeling

Correlation, between the corrupted video packets in the same frame and the distortion due to the MB mismatch in adjacent video frames, is quantified by the spatial and temporal error propagation terms in (2).

The temporal propagation loss, $f_{tp}^{i,j}$, represents the propagation of corrupted information through predictive coding. An Adaptive Intra Refresh (AIR) algorithm, which uses a selected number of intra coded MBs in a video frame, is used to prevent the error propagation. The temporal error propagation is calculated at MB level. Let $f_{tp,mb}^{k,j}$ be the temporal error propagation of k^{th} MB in j^{th} frame. $f_{tp,mb}^{k,j}$ depends on the coding mode used in the encoding of the MB. $f_{tp,mb}^{k,j}$ is calculated for inter coded MBs as

$$f_{tp,mb}^{k,j} = P_{TP}^j [\rho_{u,pv}^{i,j} \cdot E(D_{s_con,mb}^{k,i,j}) + \rho_{d,pv}^{i,j} \cdot E(D_{t_con,mb}^{k,i,j}) + f_{sp,mb}^{k,j} + (1 - \rho_{u,pv}^{i,j}) \cdot f_{tp,mb}^{k,j-1}] \quad (8)$$

and for intra coded MBs as

$$f_{tp,mb}^{k,j} = P_{TP}^j [\rho_{u,pv}^{i,j} \cdot E(D_{s_con,mb}^{k,i,j}) + f_{sp,mb}^{k,j}] \quad (9)$$

where $E(D_{s_con,mb}^{k,i,j})$ and $E(D_{t_con,mb}^{k,i,j})$ represent the spatial concealment distortion and temporal concealment distortion of the MB respectively. As before, i and j represent the i^{th} video packet in j^{th} video frame. $f_{sp,mb}^{k,j}$ is the spatial error propagation of the k^{th} MB in j^{th} video frame. P_{TP}^j quantifies the fraction of distortion of the Macro Block, which should be considered in the propagation loss calculation. The P_{TP}^j is a function of effective channel bit error rate, ρ_{beff} , and experimentally it is found to be approximated to

$$P_{TP}^j = (1 - (1 - \rho_{beff})^{F^j}) \quad (10)$$

$$\rho_{beff} = (X \cdot \rho_{b1} + Y \cdot \rho_{b2}) / (X + Y) \quad (11)$$

F^j is size of the j^{th} frame. X and Y are the total amount of data belonging in the first partition and the second partition respectively. $f_{tp}^{i,j}$ is computed as

$$f_{tp}^{i,j} = (1 - P_{u,pv}^{i,j}) \cdot \sum_{k \in \Omega} f_{tp,mb}^{k,j-1} \quad (12)$$

Ω denotes the sets of macro blocks in the i^{th} video packet in the j^{th} video frame. The spatial error propagation terms, $f_{sp}^{i,j}$ and $f_{sp,mb}^{k,j}$ depend on the number of corrupted video packets in a frame and only propagate through the spatial concealment process. The $f_{sp}^{i,j}$ and $f_{sp,mb}^{k,j}$ are modeled as,

$$f_{sp,mb}^{k,j} = P_{SP}^j \cdot \rho_{u,pv}^{i,j} \cdot E(D_{s_con,mb}^{k,i,j}) \quad (13)$$

$$f_{sp}^{i,j} = P_{SP}^j \cdot \rho_{u,pv}^{i,j} \cdot E(D_{s_con,pv}^{i,j}) \quad (14)$$

$$P_{SP}^j = N_{vp}^j \cdot (1 - (1 - \rho_b)^{VP^j}) \quad (15)$$

where N_{vp}^j and VP^j represent the number of video packets and the average size of a video packet in the j^{th} frame respectively.

III. ENERGY OPTIMIZED UEP SCHEME

Let the user requested video quality in terms of average frame PSNR be Q_{target} . Total channel interference and the noise experienced is denoted by noise power spectral density, N_o . The minimum required transmission energy for an information bit to satisfy the user quality requirement under a given channel condition is $E_{b,min}$. The expected video frame quality, $E(Q_{f,min})$, is computed using (1). It is assumed that the data on both higher priority and lower priority channels are transmitted with equal bit energies, $E_{b,min}$. Point A in Fig. 3 represents the estimated quality, $E(Q_{f,min})$. Point E shows the expected quality, if the transmission energy on both channels is incremented by 1 dB.

The goal is to find the combination of transmission energies in the two channels, which maximize the current video frame quality for an increment in the transmission energy. The corresponding point for optimum energy allocation is to be found in the vicinity of Point B. The possible combinations of transmission energy allocation in different priority channels are listed in Table I. Even though, it is valid to consider other transmission energy levels, the settings listed in Table I are more likely to provide optimum energy allocation. Moreover, this simplifies the searching procedure. The expected video frame qualities are computed for the energy settings listed in Table I and the corresponding points are shown in Fig. 3.

TABLE I. POSSIBLE COMBINATION OF TRANSMISSION ENERGY ALLOCATION.

point	Energy level on high priority channel	Energy level on low priority channel	point	Energy level on high priority channel	Energy level on low priority channel
A	$E_{b \min}$	$E_{b \min}$	F	$E_{b \min} + 1$	$E_{b \min} + 2$
B	$E_{b \min}$	$E_{b \min} + 1$	G	$E_{b \min} + 2$	$E_{b \min}$
C	$E_{b \min}$	$E_{b \min} + 2$	H	$E_{b \min} + 2$	$E_{b \min} + 1$
D	$E_{b \min} + 1$	$E_{b \min}$	K	$E_{b \min} + 2$	$E_{b \min} + 2$
E	$E_{b \min} + 1$	$E_{b \min} + 1$			

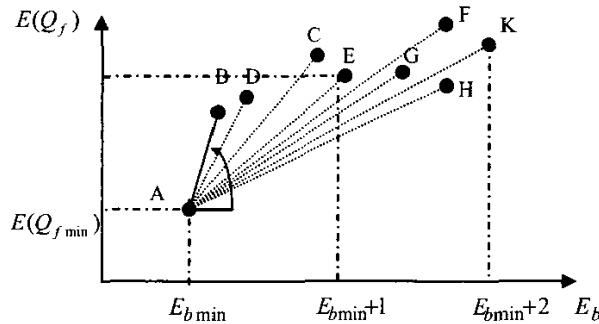


Figure 3. Calculation of optimal transmission energy levels.

The transmission energy levels corresponding to the point that shows the highest gradient from point A provide the optimal energy levels for the transmission of the current video frame.

This algorithm operates at the video frame level to find the optimal operating point. It guarantees the end user quality requirement and the optimal energy setting throughout the transmission.

IV. SIMULATION SET-UP

Realization of the proposed energy optimized UEP algorithm over UMTS network is shown in Fig. 4. At the encoder, each encoded video frame is separated into two streams based on MPEG-4 data partitioning. The separated streams are mapped on to two transport channels. The high priority data is sent over the highly protected channel, which is protected with a 1/3 rate convolutional code. A 1/2 rate convolutional code is used to protect the lower priority channel. At the physical layer, the information on the transport channels is allocated the selected transmission bit energy and is multiplexed on to the same physical channel for the transmission over the air interface.

A UMTS physical link layer simulator is developed. The simulator includes all the radio configurations, channel structures, channel coding/decoding, spreading/de-spreading, modulation parameters, transmission modeling and their corresponding data rates according to the UMTS specifications. The transmitted signal is subjected to a multi-path fast fading environment, where the power-delay profiles are specified in [8].

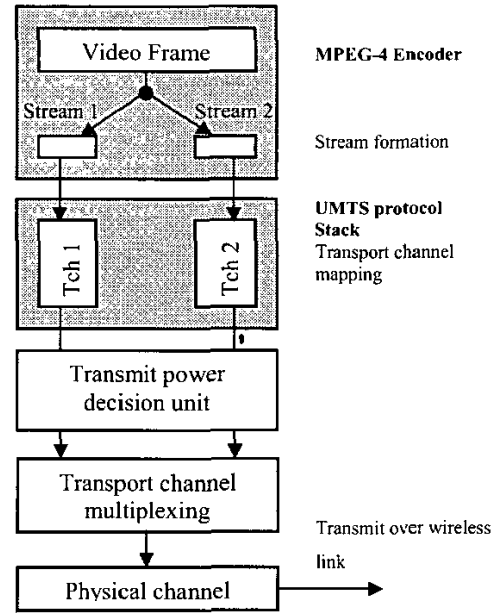


Figure 4. Realisation of proposed UEP over UMTS.

TABLE II. CHANNEL BLER FOR VEHICULAR A ENVIRONMENT.

E_b/N_0	1/2 CC	1/3 CC	E_b/N_0	1/2 CC	1/3 CC
3 dB	0.92	0.78	8 dB	0.047	0.013
4 dB	0.78	0.53	10 dB	0.0020	0.0010
6 dB	0.31	0.13	12 dB	0.0010	0.000

CC - convolutional code

The multipath-induced inter symbol interference is implicit in the developed chip level simulator. A detailed discussion of the channel simulations can be found in [9]. Using the developed simulator, error characteristics of transmission channel are simulated for a range of channel condition and for different physical layer configurations.

The Vehicular A propagation condition and down link transmission is considered in the experiment discussed in this paper. The mobile speed is set to 50 km/h. A spreading factor of 32 is used in the physical channel configuration. The experimentally evaluated channel Block Error Rates (BLER) over the Vehicular A environment are listed in Table II.

Video sequences are encoded according to the MPEG-4 simple profile [7] format. This includes the error resilience tools such as video packetisation, data partitioning, and reversible variable length code. The first video frame is Intra coded while others use Inter coding. TM5 rate control algorithm is used to achieve a smoother output bit rate, while an adaptive intra refresh algorithm [3] is used to stop temporal error propagation. ITU test sequence "Suzie" is used as the source signals in the experiments. The QCIF (176 by 144 pixels) sequence is coded at 10 fps.

Use of spreading factor 32 in channel configuration permits 64 kbps and 97 kbps information rates with 1/3 rate and 1/2 rate convolutional coding respectively.

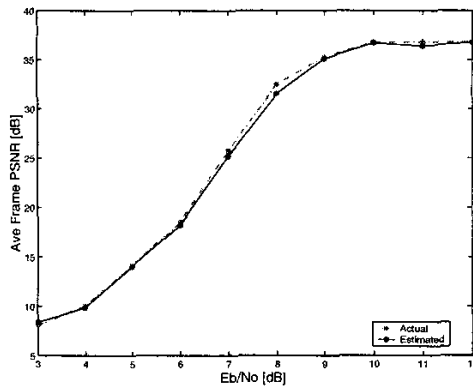


Figure 5. Validation of model performance for DP based UEP scheme.

For the proposed algorithm, video coded at 88 kbps provides the appropriate source-channel coding ratio for the "Suzie" sequence.

V. RESULT AND ANALYSIS

The accuracy of the developed distortion model is evaluated by comparing the estimated performance and the actual video performance over a simulated UMTS environment. Video performance is measured in terms of frame Peak Signal to Noise Ratio (PSNR), which is the standard objective quality measurement. Each experiment was repeated 20 times in order to average the effect of bursty channel errors on the performance. The average frame PSNR is obtained by averaging over 6000 frames and the results are shown in Fig. 5. The experimentally evaluated channel BLER (see Table II) is used in the theoretical performance calculation. As can be seen in Fig. 53, the theoretical PSNR values closely match the actual PSNR values for a wide range of channel conditions.

Experiments were conducted for a range of channel conditions for both a Data Partition based UEP (DP based UEP) scheme and the proposed power optimized UEP scheme. Results are shown in terms of average E_b/N_o vs average frame PSNR in Fig. 6. The performances of 1/2 rate and 1/3 rate Convolutional Code without application of UEP are also shown. The figure clearly demonstrates that efficient energy utilization is achieved with the proposed method compared to the traditional DP based UEP scheme as well as the non-UEP schemes. Video performances are limited by the quantization distortion with good channel conditions. In such situation, increasing the transmit power will not further increase the performance. This effect is well captured by the proposed algorithm and the highest allocated transmit E_b/N_o is limited to 11.2 dB. For the allocation of higher transmission energies, the proposed algorithm shows close performance to that of DP based UEP scheme. However, proposed algorithm considerably out-perform the DP based UEP scheme at lower transmit energy. For example, for transmission of "Suzie" sequence, the achieved average frame PSNR with DP based UEP scheme is 18 dB for 6 dB energy allocation. The proposed

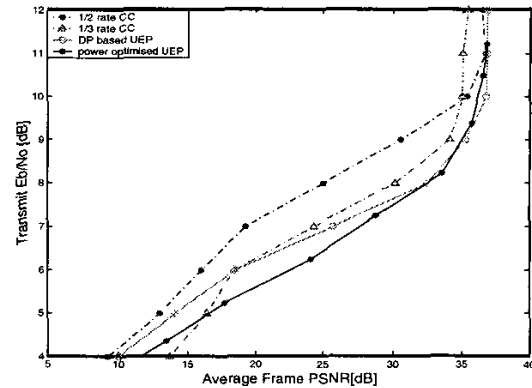


Figure 6. Performance of proposed algorithm for transmission of "Suzie".

power optimized UEP scheme results in average frame PSNR of 24 dB for the same transmit bit energy allocation.

VI. CONCLUSION

An energy efficient network compatible performance enhancement method is proposed for video communication over direct-sequence CDMA cellular networks. Prioritized video information is transmitted over two different transport channels with different error protection capabilities. Transmit energy for each bearer is selected to maximize the expected frame quality for an increment in transmit energy. The experiment that is carried out over the simulated UMTS system shows significant performance improvement with the proposed algorithm compared to the traditional DP based UEP schemes and non-UEP schemes.

REFERENCES

- [1] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol: 12 Issue: 6, June 2002 pp: 411-424.
- [2] Il-Min Kim And Hyung-Myung Kim; "An optimum power management scheme for wireless video service in CDMA systems", *IEEE Trans. on Wireless Comm.*, Vol: 2 Issue: 1, Jan. 2003.
- [3] ISO/IEC JTC 1/SC 29/WG 11, "Information technology - generic coding of audio-visual objects - part 2: Visual", *ISO/IEC 14496-2*: 2001, July 2001.
- [4] 3GPP TS 25.214, "Physical layer procedures (FDD) Release 5", V5.5.0 (2003-06).
- [5] R. Talluri, "Error-resilient video coding in the ISO MPEG-4 standard", *IEEE Communications Magazine*, pp. 112-119, June 1998.
- [6] S. Worrall, S. Fabri, A. Sadka, and A. Kondo, "Prioritisation of Data Partitioned MPEG-4 over Mobile Networks" *ETT - European Trans. on Telecom.*, V. 12, No. 3, May/June 2001.
- [7] 3GPP TS 26.110, "Codec for circuit switched multimedia telephony service; General description (Release 5)", V5.0.0 (2002-06).
- [8] 3GPP TR 101 112, 1998, "Selection procedures for the choice of radio transmission technologies of the UMTS", v.3.2.0.
- [9] C. Kodikara, S. Worrall, S. Fabri, and A. Kondo, "Performance evaluation of MPEG-4 video telephony over UMTS", *3G2003*, 25-27 June 2003, London, UK, pp. 73-77.